

AD-A096 153

ARMY DUGWAY PROVING GROUND UT
HANDBOOK FOR EVALUATING ECOLOGICAL EFFECTS OF POLLUTION AT DARC--ETC(U)

F/G 5/2

DEC 79 C F PINKHAM, D A GAUTHIER

UNCLASSIFIED DPG-HB-80-807

NL

304
Spectra

END
DATE
4-2-79
DTIC

AD A 096153



AD _____
1
2
R S

RDTE Project No. _____

TECOM Project No. 0-CO-520-EOC-004 ✓

DPG Document No. DPG HB 80-807 ✓

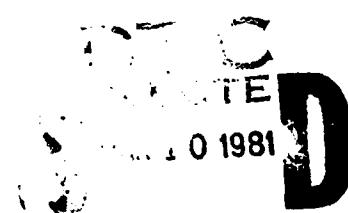
REVIEWED

HANDBOOK

FOR EVALUATING ECOLOGICAL EFFECTS OF POLLUTION
AT DARCOM INSTALLATIONS

VOLUME 7

HANDLING DATA



DECEMBER 1979

A

U.S. ARMY DUGWAY PROVING GROUND
Dugway, Utah 84022

Supervised
by [signature]

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

DRC FILE COPY

81 3 09 025

DISPOSITION INSTRUCTIONS

DESTROY THIS DOCUMENT WHEN NO LONGER NEEDED
DO NOT RETURN IT TO THE ORIGINATOR

DISCLAIMER

THE FINDINGS IN THIS DOCUMENT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS. THE USE OF TRADE NAMES IN THIS REPORT DOES NOT CONSTITUTE AN OFFICIAL ENDORSEMENT OR APPROVAL OF THE USE OF SUCH COMMERCIAL HARDWARE OR SOFTWARE. THIS REPORT MAY NOT BE CITED FOR PURPOSES OF ADVERTISEMENT.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Item 20 (con't)

The handbook covers the following areas in seven volumes of which this is Volume 7: (1) basic questions that need answering, (2) conducting the preliminary investigation of the problem, (3) determining the specific effects of a pollutant (the first three volumes are essentially library efforts), (4) terrestrial sampling, (5) aquatic sampling, (6) unexpected declines in animal populations and (7) handling data.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

TABLE OF CONTENTS

	<u>Page</u>
CONCEPT	ii
USE OF THIS VOLUME	iii
FOREWORD	iv
STATISTICAL TERMINOLOGY	1
 <u>STATISTICAL METHODS -</u> 	
INTRODUCTION	10
UNIVARIATE PROCEDURES	12
MULTIVARIATE PROCEDURES	18
FINITE POPULATION SAMPLING	22
COMMUNITY ANALYSIS	22
 <u>APPENDIXES</u> 	
A. LIST OF TESTS	A-1
B. DISTRIBUTION LIST	B-1

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Margin of Error as a Function of Sample Sizes up to 1000 in a Finite Population	21
2. Margin of Error as a Function of Sample Sizes up to 100 in a Finite Population	22

CONCEPT

The use of appropriate statistical techniques is essential to the success of a pollution ecology survey (PES). From the commander's and the environmental scientist's viewpoint, success is: (1) results that will withstand critical examination, (2) methods that produce these results with the least expenditure of time and money, (3) the identification of differences between samples that would otherwise be lost in reams of data and (4) a tool for understanding how the results were arrived at and the meaning of these results. Used properly, this volume will help the PES team satisfy the above definition of success.

Because the design of the sampling program is a complicated subject, a statistician should always be consulted or made available to the team. What follows will give an appreciation of the wide range of statistical methods available to the PES team and an understanding of the conditions under which each test is appropriate. However, it is not intended to be all-inclusive, but rather to discuss the methods that will be used commonly in the analysis of the data. The examples used are oriented toward ecological problems to make them more meaningful to the reader.

USE OF THIS VOLUME

This volume introduces the non-statistically oriented environmental scientists and interested paraecologists to the statistics commonly involved in a PES. It has been made as simple as possible by eliminating mathematical theory.

The first section begins with basic terminology and gradually builds upon prior definitions to more complex concepts. The remaining sections identify the conditions under which a given test are appropriate. Footnotes provide a reference to the computer program incorporating a particular test.

During Phase IV (design of the sampling and analysis program - page vii, Volume 1), each environmental scientist should develop an appreciation for the concepts that are outlined in this volume. Armed with these concepts, he can then talk intelligently with the statistician during their joint endeavor to identify the most appropriate statistical design for the remainder of the survey.

During Phase V (sampling program - page viii, Volume 1), the paraecologist will use this volume to refer to the computer programs for the statistical tests identified in phase IV. The instructions for the programs will show the paraecologist how to record the data or transfer the data from the log so that it can be easily entered in the computer.

During Phase VI (analysis of data - page viii, Volume 1), the paraecologist will give the data to key-punch operators or enter the data himself at a keyboard or through some other input device. He will check the entered data for errors and submit the verified data to automated data processing (ADP) for analysis using the designated program. The paraecologist may also be directed by the environmental scientists to organize the computer output prior to phase VII (interpretation of the data analysis - page viii, Volume 1) which is conducted by the environmental scientists.

FOREWORD

This volume was revised by David A. Gauthier and Dr. Carlos F. A. Pinkham.

This volume was prepared by Dr. Robert L. Stearman, U.S. Army Dugway Proving Ground, Dugway, UT 84022.

Gary L. Sutton, U.S. Army Dugway Proving Ground, Dugway, UT 84022, contributed written material to Volume 7.

In addition, the following people have contributed information used in Volume 7:

Dr. Victor C. Beal, Jr., Veterinary Services, U.S. Department of Agriculture, Hyattsville, MD 20783

Paul E. Carlson, U.S. Army Dugway Proving Ground, Dugway, UT 84022

Russell Jensen, Computer Center, University of Utah, Salt Lake City, UT 84112

Dennis Fuller, U.S. Army Dugway Proving Ground, Dugway, UT 84022

NOTE

The pronoun "he" is used in this volume as an impersonal pronoun which encompasses he and she and has no intent of personal reference or connection.

NOTES

STATISTICAL TERMINOLOGY¹

Every field of endeavor has its own particular terminology. The terms may originate within the field or they may be borrowed from another. In all cases, however, the terms take on very exact and specific meanings within a single field and may bear only partial resemblance to the meanings of the same words when used in another field. This section is devoted to defining some of the basic statistical terms used in pollution ecology surveys (PESs).

The basic unit in statistics is the observation. An observation is a measurement. It may be qualitative (identification of kind), such as the presence or absence of a given species in a sample, or it may be quantitative (measured amount), such as the blood level of a pollutant in a given animal. The observation may be taken on a single entity, or it may be taken on several entities, such as the presence or absence of 20 given species in a sample, or the average blood level of a pollutant for the animals in a fixed area.

A population (or universe) is a group of observations that is made up of a very specific collection of measurements which is defined by the experiment or study being undertaken. For example, the population may be the concentration of the pollutant in a given species of plant in the contaminated area. The interest of the PES team and the purpose of the study will define the population.

A sample is a group of observations drawn from a population and is used to characterize the population. The population from which the sample is drawn is called the parent population. The population must be defined before the sample is drawn so only the population of interest is sampled.

The number of observations in a sample is referred to as the sample size. The size of a sample is limited only by the size of the parent population and may consist of only one observation (a sample size of one) or may encompass the entire population (complete sampling).

The type of sample which is preferred and which is required for most statistical tests and estimation procedures is the random sample. This is a sample drawn by any method which gives each and

¹For a more detailed discussion of statistical terminology, refer to: Sokal, R.R. and F.J. Rohlf, Biometry, W.H. Freeman and Company, San Francisco, 1969.

every member of the parent population an equal chance of appearing in the sample and which is such that the choice of any particular member for inclusion in the sample does not affect the chance of any other member appearing in the sample.

A concept basic to the discussion to follow is that of a population of all possible samples of a given size. To illustrate this concept, consider a population which consists of only five observations, numbered one through five for convenience, from which a sample of size three is to be drawn. The list of possible samples would include:

1,2,3
1,2,4
1,2,5
1,3,4
1,3,5
1,4,5
2,3,4
2,3,5
2,4,5
3,4,5

There are thus 10 different samples containing three of the five possible observations. These represent all of the possible different combinations of three observations which can be drawn from the given parent population. Each member of the population appears in exactly six of the 10 possible samples.

In the preceding example, the order in which the observations were drawn was not specified. Each of the 10 samples could have been drawn in six different ways. Thus, the first sample could have been drawn in any of the six following orders:

1,2,3
1,3,2
2,1,3
2,3,1
3,1,2
3,2,1

When the order is considered, there are 60 possible samples and each member will appear in 36 of the possible samples. The proportion of samples in which any member of the parent population appears, remains unchanged regardless of whether order is considered.

Now consider a hypothetical example of a sample of four snails from a population in which 50% were coiled to the left (left-handed) and 50%

were coiled to the right (right-handed). If order is not considered, there are five possible outcomes:

four left-handed snails
three left-handed and one right-handed snails
two left-handed and two right-handed snails
one left-handed and three right-handed snails
four right-handed snails

This listing would seem to imply that an outcome of four left-handed snails would have the same chance of occurring as two left-handed (L) and two right-handed (R) snails. There are 16 possible samples of size four when order is considered:

		LLRR
		LRLR
LLLR	LRRL	LRRR
LLRL	RLLR	RLRR
LRLL	RLRL	RRLR
LLLL	RLLL	RRLL
		RRRR

Thus, the chance of getting two left-handed and two right-handed snails is six times that of getting four left-handed snails when order is considered.¹ As a rule of thumb, it is best to consider order of draw until it can be shown to have no effect.

The distribution of a population is the relationship between the measurements and their frequency of occurrence. In the example of left and right-handed snails, the 16 possible samples were tabulated so that the distribution or underlying structure of the population could be visualized. Tables and graphs are the usual means of portraying the distribution of any population. There are numerous ways in which populations can be distributed. The Gaussian or normal distribution is frequently encountered in statistical studies. Such distribution resembles a vertical cross-section of a bell and is, therefore, frequently referred to as a bell-shaped distribution, or more simply a bell curve.

There are many characteristics of distributions which might be used to differentiate between populations. Among the more frequently used are mode, percentile, median, range and limits:

The mode is the measurement with the greatest frequency of occurrence. In a frequency curve of the distribution, this is shown as a peak or hump in the curve. Distributions with one mode are referred to as unimodal while those with two modes are referred to as bimodal.

¹As well it might be in a PES. For example, suppose pollution were to alter the ratio of left-handed to right-handed snails.

A percentile is that value below which a given portion of the distribution of observations occur. Thus, the 10th percentile would be the value below which 10 percent of the distribution of observations will fall.

The median of a population is a measurement which divides the distribution into two parts, each having the same number of observations, one which consists of measurements less than the median and the other consisting of measurements greater than the median.

The range of a set of measurements is the difference between the largest and smallest values in the set. That is, it is the difference obtained by subtracting the smallest measurement from the largest. The limits of a population are the smallest and largest measurements. That is, a population with limits of 1 and 100 has a range of 99.

Mathematical functions of the observations in a population can also be utilized to differentiate between populations. A parameter is a numerical constant appearing in, or derived from, the mathematical equation for a distribution. This constant takes on a particular value for a given population but may vary from population to population. A parameter, then, is a characteristic of a population utilized in mathematically defining the distribution of that population. There are many parameters. Each class of distributions has its own parameters which are used in characterizing the various members within the class. However, attention will be confined to two parameters used in characterizing the normal distribution; the mean and the variance.

The mean of a population is the arithmetic average of the observation it contains and is found by adding the observations of the population together, followed by division of the sum by the number of observations in the population. The mean is a centering parameter, giving a measure of the location of the distribution.

The population variance is the parameter measuring the spread, dispersion or variability of the measurements in a population. It is defined as the arithmetic average of the squared deviations¹ from the population mean of the measurements making up the population. The population mean is subtracted from each individual observation to obtain the deviation; these deviations are then squared and added together, followed by division of the resulting sum by the number of observations in the population. The population standard deviation is the square root of the variance. The population standard deviation has the advantage of being in the same units of measure as those for

¹A deviation in this sense is the difference obtained by subtracting the observation from the mean. Squaring this difference eliminates negative values.

the population observations and the population mean. Thus, if the observations are measured in milligrams, the population mean and the population standard deviation will be expressed in milligrams.

Populations are characterized by parameters; samples are characterized by statistics. A statistic is a characteristic of a sample, a function of the observations in the sample. The sample statistics are the tools used to characterize the parent population through study of the observations in a sample. The usefulness of any particular sample statistic will depend on its ability to estimate a parameter of the parent population effectively. A concept required for further discussion of how well a statistic estimates a parameter is that of the estimator and an estimate.

An estimator is the arithmetic procedure used to obtain a statistic from the observations in a sample. An estimate is the actual number or value obtained when this procedure is applied to a particular sample. An estimator, being the procedure used, will remain the same from sample to sample. But estimates, being the actual values obtained, will differ from sample to sample as long as there is any variation among the observations in the parent population.

The expected value for a function of the measurements in a population is the average or mean value of that function for all the measurements in the population. Hence, if the function consists of the measurements themselves, the expected value would be the mean, since both are defined as the average of the measurements. If the function is the square of the deviation of the measurement from the population mean, the expected value is the variance of the population.

The concept of an expected value can be applied to that of an estimator, utilizing the concept of the population of all possible samples of a given size. If the estimator is applied to each sample in this population of samples, a population of estimates is obtained. If the expected value of the population of estimates is equal to the parameter being estimated, then the estimator, or method of estimation, is said to be unbiased. If the expected value is not equal to the parameter, then the estimator is said to be biased and the difference between the two is called the bias. If the parameter is less than the expected value of the population of estimates, the estimator is said to be positively biased while if the parameter exceeds the expected value, the estimator is said to be negatively biased.

Bias is only one of the criteria used for judging how well a statistic estimates a parameter. It is, however, a main one.

The sample mean is defined as the arithmetic average of the observations in a sample and is obtained by adding the observations of the sample together, followed by the division of the sum by the number

of observations in the sample. The sample mean is an unbiased estimator of the population mean.

The statistic used to estimate the population variance is the sample estimate of the variance which is found by squaring the deviations of the observations from the sample mean, adding them together, (sums of squares) and then dividing the resulting sum by one less than the number of observations in the sample. One less than the sample size is used rather than the sample size itself in order to provide an unbiased estimator of the population variance¹. The divisor, one less than the sample size, is called the number of degrees of freedom of the sample estimate of the variance. The square root of the sample estimate of the variance is the sample estimate of the standard deviation of the population.

Significance tests are an essential part of statistical procedures. An attempt will be made here, with the help of a simple example, to provide an understanding of the structure underlying these tests and the terminology for establishing the criteria used.

Pollution example: Two species are affected equally by a pollutant. This is our hypothesis (theory or proposition) to be tested. This particular hypothesis is called the null hypothesis because it states there is no difference in the sensitivity of the two species in the population sampled. Assume that the sample taken indicated that the first species was twice as sensitive as the second. In one case, we may decide not to reject the null hypothesis because a doubling of the sensitivity is not abnormal. In another case, we may decide that the null hypothesis should be rejected because so deviant a sample normally would not be expected from a population of two equally sensitive species. Two possibilities exist: (1) The two species are equally sensitive or (2) they are not equally sensitive. In both situations, two decisions are possible: (a) not rejecting the null hypothesis or (b) to reject the null hypothesis. The combination (1) (a) gives a correct decision, no error is made. The combination (1) (b) gives an incorrect decision, and a type I error (the rejection of a true null hypothesis) is made. The combination (2) (a) also gives an incorrect decision and a type II error (not rejecting a false null hypothesis) is committed. The combination (2) (b) is a correct decision and, again, no error is made.

¹On an average, the sample variance underestimates the magnitude of the population variance. In order to overcome the bias, sums of squares are divided by one less than the sample size ($n-1$) rather than the sample size (n); so that the sample variances will be unbiased estimators of the population variance.

An important step in statistical tests is to decide what magnitude of type I error can be tolerated. Generally, scientists are willing to accept the commission of a type I error in 5 percent of the samples. That is five out of every 100 samples taken from a given population would be so deviant as to appear to be from another population. When the type I error is expressed as a percentage or decimal; it is known as the significance level. Thus a type I error of five out of 100 samples is a significance level of 5 percent or 0.05.¹ The area in the tails (extremes) of the frequency distribution curve representing the significance level (in this example, the five most deviant samples out of every 100) is called the rejection region or critical region of a test. The remainder of the area under the curve is called the non-rejection region.

Rejecting fewer of the deviant samples (one out of 100 instead of five out of 100) from the population reflected in the null hypothesis leads to the commission of a type I error one percent of the time. This would seem to be a better approach except that now we are more apt to make a type II error.

It seems we are caught between a rock (committing a type I error) and a hard place (committing a type II error). There is a way out, however - a way that is generally inadequately understood and one which involves two aspects of experimental error.

(1) Obviously, it makes no sense to look for a difference that is less than the precision of the measurement (e.g. if we are measuring the length of fish to the nearest centimeter, a meaningful difference between two populations must be greater than one centimeter). Other factors enter into measurement precision, such as human and equipment error. These can only be quantified through experience and must be determined by the environmental scientists on the team.

(2) Also, of importance in identifying a meaningful difference, is a foreknowledge of what might be termed the "biological variability". For example, when conducting a PES, we would want to ignore the difference between aquatic samples caused by natural variations such as light and temperature, but detect the difference caused by various levels of pollution. Ideally, experiments will have been performed to establish biological variability. In almost every case, this important datum is not available and the environmental scientist must make a best estimate of biological variability based upon his experience.²

¹This is known as the α level.

²He also should work with the statistician to minimize the influence of biological variability.

Once we have established a meaningful difference, the choice of the significance level of a test rests on balancing the chances of type I and type II errors and the impact that sample size has on the type II error.

If a large difference is meaningful, then the chance of making a type II error is (likely) small. On the other hand, if a small difference is meaningful, then the chance of making a type II error is (likely) large. When a small difference is meaningful, then, the obvious question becomes: "how do we decrease the chance of making a type II error?" The answer lies in why we used the term "(likely)" above. The variance of the population made up of sample means, obtained from the population of all possible samples of a given size, is equal to the variance of the parent population divided by the sample size. Thus, the larger the sample size, the smaller the variance of the distribution of sample values. Therefore, if the sample size is too large, we likely will reject a null hypothesis which is false, even though the true parameter value and that of the null hypothesis may not be sufficiently different to be meaningful. On the other hand, if the sample size is too small, we may fail to reject a null hypothesis which is false even though the true parameter value and that given by the null hypothesis are sufficiently divergent to be meaningful. The trick, then, is to select a sample size that will detect a meaningful difference. Because this step is somewhat complicated the team statistician must be consulted to determine the proper sample size.

Once the sample size is determined, we must ask ourselves, do we have the resources to obtain the required sample size? If not, we will have to accept a larger risk of making a type II error; we will have to look for differences larger than the meaningful difference; or we will have to abandon the study.

In the pollution example given on page 6, the alternative to the null hypothesis is that the two species are not equally sensitive. This is called a two-tailed test because both tails (extremes) of the distribution will be considered. That is, the null hypothesis will be rejected if the first species is more sensitive than the second or the second is more sensitive than the first. Had we established the alternative hypothesis that the first species was more sensitive than the second, then a one-tailed test would be indicated because we are interested in only one tail of the distribution.

In summary, the steps in a statistical significance test may be stated in question form:

- (1) What is the null hypothesis?

- (2) What are the "interesting" alternatives to the hypothesis?
(This determines whether the test is one-tailed or two-tailed.)
- (3) What chance of a type I error will be allowed? (This sets the significance level for the test.)
- (4) What chance of a type II error will be allowed? (This establishes the desired sample size.)
- (5) Is the sample size realistic?
- (6) Given the above, what is the critical region for the test?
- (7) What is the sample value obtained from the data?
- (8) Does the sample value fall into the critical region for the test? (If the sample value falls into the critical region, reject the hypothesis and state which alternative to the null hypothesis appears to be true. If the sample value does not fall into the critical region, do not reject the hypothesis.)

When a sample value falls into the critical region of the test, it is said to be significantly different, statistically, from the value given by the hypothesis. If the sample value does not fall into the critical region, it is said to be not significantly different, statistically, from the value given by the hypothesis.

STATISTICAL METHODS¹

INTRODUCTION

Pollution Ecology Surveys require experimental and reference study areas. An experimental study area is the area where the polluting activity is suspected of causing environmental damage. A reference study area is an area not affected by the polluting activity, and one in which all relevant variables (climate, biota, substrate and topography) are as close as possible to those in the experimental study area.

For past and proposed potentially polluting activities, the reference and experimental areas are often the same. The area is studied before the operation of a polluting activity to establish a baseline², and concurrently during the operation to detect changes from the baseline that indicate environmental stress. Usually, time is considered an irrelevant variable.³

In cases involving present polluting activities, the experimental area clearly cannot double as the reference area unless some of the techniques for studying past polluting activities can be applied (refer to Volume 4). When they cannot, a reference area is chosen that most closely resembles the experimental area. There is an inherent risk in this procedure because there may be subtle microhabitat differences that can have a profound effect on the biological indicator. Furthermore, areas regarded as appropriate references could have a habitat similar to that of the experimental area because they are exposed to similar substances. These potential pitfalls may be overcome by carefully tailoring the selection of a reference area, to the unique combination of conditions for each polluting activity. However, it is highly preferable that a study on a proposed action be initiated a year or more before, rather than after a planned activity commences, because a reference area at the same site as the experimental area is more desirable.

¹In the event of an unexpected decline in animal populations, a good reference is: Veterinary Services, U.S. Department of Agriculture, Hyattsville, MD. Regulatory statistics, 5th Ed., V.C. Beal Jr., 1977.

²The baseline can include species structure (numbers and kinds of species present) and levels of pollutants in the species sampled.

³This may not be so, for example, if relevant variables in the study/experimental area are known to undergo cycles which may in turn affect the biological indicators.

Individual sample sites are located within the experimental and reference areas. The exact location of the sample site is determined by a random selection system.¹

The field of statistics offers an extremely large variety of procedures, far more than would ever be utilized in any given PES. The choice of the methods to be employed depends upon the particular pollution problem and the population or populations defined by that problem. To begin with, usual statistical theory assumes a normal distribution of the measurements in the population, giving rise to a number of parametric statistical methods. On the other hand, occasions arise when normal distribution of the measurements cannot be assumed or is known to be false, in which case non-parametric procedures are used.² Both parametric and non-parametric techniques will be mentioned.

Another division in statistical procedures is based on the number of variables³ being tested simultaneously or the number of dependent⁴ variables being considered. The older statistical methods are essentially univariate in nature, that is they only test one variable at a time or consider one dependent variable. Thus, if in a field study of a particular type of rodent, we were interested only in testing level of the pollutant in the blood (the variable of interest), then our tests would be univariate even though we might want to compare the average concentration among two or more populations (sources). The problem remains univariate even when we become interested in considering the dependence of the blood level on the distance of the burrow from the source of pollution, the strength of the source, the concentration of the pollutant in each of several potential foods, etc., because we would have one dependent variable and relate it to several independent variables. The univariate nature of the problem is still there even if our interest broadens to several variables such as pollutant concentration in the blood, various organs, hair, urine, feces,

¹Computer Branch, Analysis and Computations Division, U.S. Army White Sands Missile Range, NM 88002. Univac 1103 Library Directory, Revision 1, Section 4.2, page 4-2, 1 May 1974.

²Ranking, a commonly used non-parametric method, consists of assigning a numerical order (i.e., in sequence) to observations from highest to lowest or visa versa. In so far as the rank consists of digital values, it ignores the relative distance between successive observations.

³A variable is the measurement of interest without regard to its source.

⁴Dependence (a characteristic of regression analysis) implies a causal relationship of the effect of the independent variable(s) on the dependent variable(s). It is not always easy to identify which are dependent or independent variables which is another reason the statistician and environmental scientists must work closely together.

etc., as long as we continue to test these other variables individually, or one at a time. The procedures become multivariate when we wish to test the effect of the pollutant on all of these variables simultaneously, that is to look at the effect of the pollutant on the total organism. Another example of a multivariate problem would arise if we were sampling several different types of animals and wished to test whether the average concentration of the pollutant in the blood differed for all of the animals simultaneously, between a control area and a contaminated one. The procedures also become multivariate when we wish to consider more than one dependent variable.

The problem is that the significance level admits to allowing a certain proportion of type I errors to each test. If only a few variables were tested, there is little threat of having made a cumulative type I error. However, if many variables were tested, the likelihood that you have made a cumulative type I error becomes very real. In other words, the significance level of the individual test has little meaning. On the other hand, with multivariate analysis, the probability of making a type I error is the significance level, because the variables are tested simultaneously. Therefore when more than several variables are being tested, or there is more than one dependent variable, multivariate techniques are appropriate.

One basic principle should be mentioned prior to exploring the various tests, namely, always use the most powerful test available. That is use the test with the best chance of differentiating between the alternatives. Under this basic principle fall the following suggestions: never use a non-parametric test when a test based on a normal distribution would apply, and never use a "shotgun" test when one with specific alternatives or a specific hypothesis can be used. With modern computers, the statistician is now equipped with test procedures which in former years would have been virtually impossible because of the large number of time-consuming calculations involved and the attendant problems of errors in number handling and computations. It now boils down to picking the best of the lot or making the test procedure fit the problem.

UNIVARIATE PROCEDURES

The simplest test among the univariate procedures is the test of the difference between two means, such as the difference in average blood level of a pollutant in a given species between a reference area and an experimental area. The normal distribution test utilized for this purpose is Student's t-test¹. Once there are more than two

¹Section 12, page 3, UNIVAC 1108 Statistical Library (STAT-PACK), UP-4041, Revision 2, Sperry Rand Corporation, 1969.

populations involved, the analysis of variance¹ is the test of choice for the normal distribution. The analysis of variance turns out to be just what the name would imply. In the analysis of variance, the total variation of the observations about the over-all mean is broken down into its parts, each of which arises from a different source. Thus, if the blood level were being studied for a pollutant in several related species of rodents and among several areas, the total variation among the measurements would be broken down into such constituents as the variation in mean level among the species, the variation in the means among the different areas, the variation in species mean with area (or vice versa), and the variability of individual animals about their species-area mean.

On the basis of the analysis of variance model used (a result of the experimental design), those variance estimates which estimate the same thing are then compared by means of Snedecor's F-test². The number of different sources of variation which can be tested is large. The number of different analyses of variance is about as unlimited as the experimental designs which give rise to them.

One of the problems with the analysis of variance is that it is a "shotgun" test. The basic hypothesis is that the population means being tested are all equal. When the hypothesis is rejected, due to a statistically significant value of F, all that is known is that the population means are not equal. Nothing is said about the pattern of deviation among the population means. One method of checking for a pattern is the "individual degrees of freedom" approach. With this procedure, the variability among the sample means is broken down into component parts each of which tests a particular hypothesis concerning the pattern of deviation among the population means. These component parts are called "orthogonal contrasts" and have to be calculated by your statistician according to specific mathematical rules in order to be applicable. Another method of determining the pattern of the variability among the population means is with the Student-Newman-Keuls test³ for comparison of means. With this procedure, the sample means are ranked and grouped to provide an estimate of the pattern of the population means. A parallel test for ratios is the Heterogeneity G-test³.

¹Section 8, pages 1-30, UNIVAC 1108 Statistical Library (STAT-PACK), UP-4041, Revision 2, Sperry Rand Corporation, 1969.

²Section 12, page 4, UNIVAC 1108 Statistical Library (STAT-PACK), UP-4041, Revision 2, Sperry Rand Corporation, 1969.

³U.S. Army Dugway Proving Ground, Dugway, UT 84022. A Computer Program for the Student-Newman-Keuls Test and the Heterogeneity G-test, C.F.A. Pinkham, In Press.

Two of the basic assumptions in an analysis of variance are that the populations of measurements have a normal distribution and that the variances of the populations are equal (homogeneity of variance). Tests of these assumptions are far more sensitive to departure than the analysis of variance so that it is somewhat like setting out to sea in a row boat to determine whether conditions were safe for the passage of an ocean liner.

Both the Student's t-test and the analysis of variance have equivalent procedures among the non-parametric tests which require either no assumptions or else assumptions other than that of the normal distribution of population measurements. See your statistician for specific tests.

Another area in univariate analysis is that of regression. Here, the problem is to determine the mathematical nature of the relationship of a dependent variable to one or more independent variables. The relationship can be visualized by plotting the dependent variable against the independent variable (for instance, the concentration of the pollutant plotted against the distance from the source). A cluster of points will result. The basic tool is called "least squares regression analysis". The exploration of this field invariably starts with simple linear regression¹ in which the point is to estimate the relationship between the dependent variable, y , and the independent variable, x , when the points appear to be clustered about a straight line. The mathematical form of such a straight line is $y = a + bx$, where b is the slope of the line (the amount y rises or falls with unit increase in x), and a is its intercept (the value of y when $x = 0$).

When the points appear to be clustered about other than a straight line, the variations are almost endless. For example, one or more variables can be transformed to make the relationship linear or linearity may be achieved by using the natural logarithm of y and fitting it to x , with the equation: $\ln y = a + bx$. Other transformations allow fitting of other exotic kinds of clusters.

Regression techniques carry with them the assumptions that the population of measurements in y for any given x have a normal distribution and that the variation of these measurements about their mean remains the same from one value of x to another. This latter assumption can be overcome by use of weighted regression.² Here, each point

¹BMDOR, page 218, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computation, Number 2, University of California Press, 1971.

²Pages 129-130 and 343, Norman, N. et al., Statistical Packages for Social Sciences (SPSS), prepared by the University of Chicago, McGraw Hill, New York, 1975.

to be fitted is weighted inversely to the amount of variability to be expected among the measured population of values of y for the given x (which is only logical, so that the more variable the point, the less weight it receives in determining where the estimated line goes). The problem here arises with attempting to obtain an estimate of the variability attached to each point. The usual way around this obstacle is to take more than one y observation for each value of x . The variability can then be estimated.

The weighted regression technique can become quite sophisticated. Probit analysis,¹ the more commonly used method for estimating biological dose-response relationships in the laboratory, is a good example. Here, the concept is to estimate the mean and variance (or standard deviation) of the normal distribution of the proportion of animals in the population responding in a particular way to each specified dose.

Another statistic connected with simple linear regression is the product-moment correlation coefficient, r ,² usually known more simply as the correlation coefficient. The correlation coefficient is a unitless measure of the linearity of response of the dependent variable, y , to the independent variable, x . A correlation coefficient of zero would indicate that the mean value of y remains the same regardless of the value of x . A correlation coefficient of plus one would signify a perfect linear relationship between y and x with a positive slope (the value of y rises with increasing values in x). A correlation coefficient of minus one also denotes a perfect linear relationship between y and x , only in this case, the value of y falls with increasing values of x and the slope is negative. The square, of the correlation coefficient (r^2) is also extremely useful in that it shows the proportion of the total variability in y which can be accounted for by the linearity of response of y to changes in x . Conversely, $(1-r^2)$ reveals the portion of the total variability in y which is present as variation of the points about the fitted line.

When the relative ranks (where the observations stand with respect to the other observations in the distribution) of the values of y and x are substituted for their measurements in the equation for the product

¹ BMD038, page 357, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computation, Number 2, University of California Press, 1971.

² BMD03D, pages 60-65, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computation, Number 2, University of California Press, 1971.

moment correlation coefficient, r_s , Spearman's rank correlation coefficient,¹ is obtained. Spearman's rank correlation coefficient provides a basis for measuring the dependence of y on x without resorting to the normal distribution assumption.

Simple linear regression can be expanded to cover the relationship of a dependent variable, y , to a series of independent variables, x_1 , x_2 , x_3 , etc. This extension of simple linear regression is called multiple regression.² As an example, the level (y) of pollutant in the blood might be related to the prevailing downwind distance (x_1) and the crosswind distance (x_2) of a rodent's burrow from the source of pollution, or a grid of north-south and east-west distances from the source might be utilized. Multiple regression has a multiple correlation coefficient,³ R , which is equivalent to the product-moment correlation coefficient, r , of simple linear regression, and which enjoys the same usefulness. As in simple linear regression, when the clusters appear to be distributed about other than a straight line, multiple regression techniques can be modified to fit an nth-order polynomial⁴ relationship between y and x by letting the various x variables be various powers of x rather than measurements of different entities (for example $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$, etc.). Further complexity can be added to the case of several x variables by measuring the impact of "interaction" among variables in estimating their relationship to the independent variable. Transformations add to the sophistication of the relationships being fitted. Models relating average concentration of a pollutant to crosswind and downwind distances from a source can be explored by means of proper utilization of multiple regression techniques. Such models can serve as a basis

¹Page 288k, Norman, N. et al., Statistical Packages for Social Sciences (SPSS), prepared by the University of Chicago, McGraw Hill, New York, 1975.

²BMD03R, page 258, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computation, Number 2, University of California Press, 1971.

³Page 331, Norman, N. et al., Statistical Packages for Social Sciences (SPSS), prepared by the University of Chicago, McGraw Hill, New York, 1975.

⁴BMD03R, page 258 and BMD02R, page 233, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computation, Number 2, University of California Press, 1971.

for establishing isopleths^{1,2,3} (lines of equal concentrations, and thus equal exposure), around a source of pollution. In this type of usage, many of the above regression techniques can come into play. It all depends on the level of complexity and sophistication desired. The principle of Ockham's Razor should apply: the simpler the better.

Stepwise multiple regression⁴ is a powerful multiple regression procedure available with the advent of computers. Different versions allow either a backward or forward procedure. In the backward process, all of the independent variables are fit to an equation that relates them, then tested and those which do not have a statistically significant contribution to the equation are removed one at a time in an ascending order of their statistical significance until all those left make a statistically significant contribution to the regression equation. In the forward procedure, the independent variable with the highest absolute correlation coefficient with the dependent variable is checked and if it makes a statistically significant contribution, is used. Values are then recomputed and the remaining variable with the highest correlation coefficient, given that the first independent variable chosen was used, is taken, and so on, one step at a time. In the event one of the independent variables chosen earlier is no longer making a statistically significant contribution, given that the other chosen independent variables are in the current regression equation, the program eliminates the unneeded independent variable and proceeds to the next step. The procedure continues until none of the remaining independent variables will add, in a statistically significant fashion, to the forecast of y (dependent variable).

¹Environmental Protection Agency, Washington, DC 20460. Air Quality Display Model, prepared by TRW System Group, Washington, DC 1969.

²Calder, K.L., "A Climatological Model for Multiple Source Urban Air Pollution," Proceedings of the 2nd Meeting of the Expert Panel on Air Pollution Modeling, NATO Committee on the Challenges of Modern Society, Paris, France, Jul 1971.

³U.S. Army Dugway Proving Ground, Dugway, UT 84022. Development of Dosage Models and Concepts, Vols 1 and 2, H E Cramer, et al, Feb 1972.

⁴BMD02R, page 233, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computation, Number 2, University of California Press, 1971.

Before ending the discussion of univariate procedures, it is well to mention a tool which combines linear and multiple regression techniques with the analysis of variance, namely the analysis of covariance.¹ In a laboratory study of a physical phenomenon, a physical scientist may spend most of the time in design and fabrication of special equipment whose purpose is to stringently control all environmental factors so that the only thing that is allowed to vary is the factor under study. In biological studies, as mentioned at the beginning of this section, every attempt must be made to ensure that the untested variables of the populations being measured are as comparable as possible so that differences in measurements may be attributed to the variable being tested, such as a possible pollution source. There are, however, always uncontrolled variables which might account for the differences observed. If such uncontrolled variables can be measured, it may be possible to correct the study results for any differences which might exist among them. As shown, the tremendous variety of regression techniques and all their modifications allow for the exploration of mathematical models for all sorts of biological and physical phenomena. This tool's utility expands even further when it is realized that the entire catalog of analysis of variance techniques represents nothing more than a modification or specialization of least squares regression. This powerful tool expands naturally into the area of multivariate analysis, again thanks to the computer.

MULTIVARIATE PROCEDURES

Multivariate procedures include the test of the difference between sets of variables for two samples, such as a reference area and an area with suspected pollution, based on Hotelling's T²-statistic² which is the square of Student's t. The canonical correlation coefficient³ is an extension of the product-moment correlation coefficient to the correlation between two sets of variables. Also available are the multivariate analysis of variance², multivariate regression⁴ and multivariate analysis of covariance². In other words, there are multivariate equivalents for many of the univariate, normal-distribution procedures given under univariate procedures.

¹BMD04V, page 525, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computation, Number 2, University of California Press, 1971.

²Pages 415-420, Norman, N., et al., Statistical Packages for Social Sciences (SPSS), prepared by the University of Chicago, McGraw Hill, New York, 1975.

³Page 515, Norman, N., et al., Statistical Packages for Social Sciences (SPSS), prepared by the University of Chicago, McGraw Hill, New York, 1975.

⁴Section UCSL 705, Sinn, J.D., MESA 97 (NYBMUL), University of Chicago Computer Center, Chicago, Undated.

There are also multivariate procedures whose primary purpose is to reduce the number of variables to be handled. The objective is to produce a "parsimonious" set of variables, that is, reduce the number of variables to the smallest possible set. One approach to parsimony is principal components analysis.¹ The concept in principal components analysis is to produce those linear combinations of correlated variables that maximize the variability of the weighted sum ($y = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$) where y equals the weighted sum and the weights, w , are functions of the relative contribution of the variables to the variability. By choosing linear combinations that maximize the variability, an attempt is made to account for as much of the total variability as possible with each set of variables; thus the name, principal components, since each constitutes a principal component of the total variability shown by the original set of variables.

Factor analysis² is an attempt to account for the correlational pattern observed in a set of variables in terms of a minimum number of factors. These factors can be thought of as latent variables which are unobservable but which, if they were known, might explain the pattern. The objective in factor analysis is to regroup the data into patterns which, hopefully, can be interpreted in a meaningful fashion. The output consists of the weights to be used in the linear combinations of factors.

Discriminant analysis³ attempts to classify or categorize an object on the basis of the measurements obtained from it. Thus, if in the study of a series of reference experimental areas a function of the measurements could be found which clearly discriminated, for future studies, between reference and experimental areas, a major objective would be attained. Discriminant analysis is a tool designed for this purpose and can be utilized to classify a given object, etc., into one of two or more populations.

¹BMD02M, pages 169-168, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computations, Number 2, University of California Press, 1971.

²BMD03M, page 169, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computations, Number 2, University of California Press, 1971.

³BMD05M, page 196, Dixon, W.J. (ed.), Biomedical Computer Programs (BMD), University of California Publications in Automatic Computation, Number 2, University of California Press, 1971.

FINITE POPULATION SAMPLING

Most of the discussion on statistics to this point has assumed an infinite population, but in dealing with flora and fauna in a restricted area, this no longer holds and more attention must be paid to finite population theory.

For the infinite population, the variance of the population of sample means, taken from the population of all possible samples of a given size, is equal to the variance of the parent population divided by the sample size. In a finite population, if the sample size is represented by n and the parent population size by N , the variance of the population of sample means is changed only by multiplying the variance for the infinite population by the factor, $(N-n)/N$, which is called the "finite population correction" (fpc). In practice, the fpc may be disregarded whenever the sampling fraction, n/N , does not exceed five percent.

The impact of the finite population correction on the estimation of sample size is shown in Figures 1 and 2, for the test of the difference between the means of two finite populations with a significance level of five percent. In these figures the margin of error¹, d , is plotted against sample size for various population sizes. The two populations are assumed to have the same size and also the same variance. Similar techniques are available for estimating sample size for other tests. The primary thing to remember is that the finite population correction should be used if the sample size exceeds five percent of the population size.

COMMUNITY ANALYSIS

In attempting to compare the structure and organization of biological communities in a PES, two basic areas are of special interest. The first compares the biological diversity (the numbers of kinds of organisms without regard to the taxa present per unit area or volume) within a community with that of another community, and the second compares the biotic similarity (the numbers of kinds of organisms with regard to the taxa present per unit area or volume) among communities.

¹The smallest difference between two statistically significantly different means that would be detectable with the given sample size, expressed as a percent of the (sample estimate of the) standard deviation.

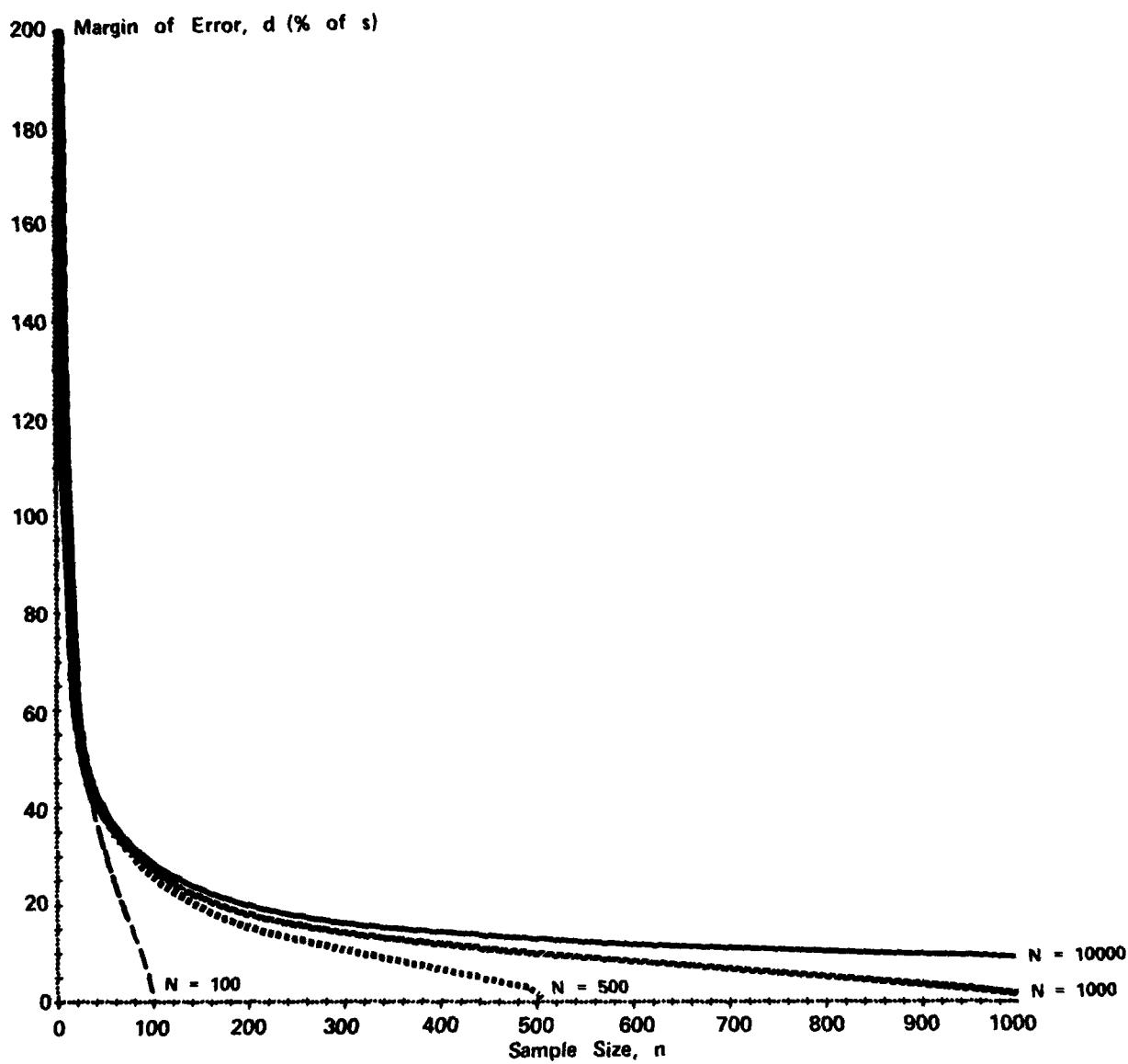


Figure 1. Margin of Error as a Function of Sample Sizes up to 1000 in a Finite Population

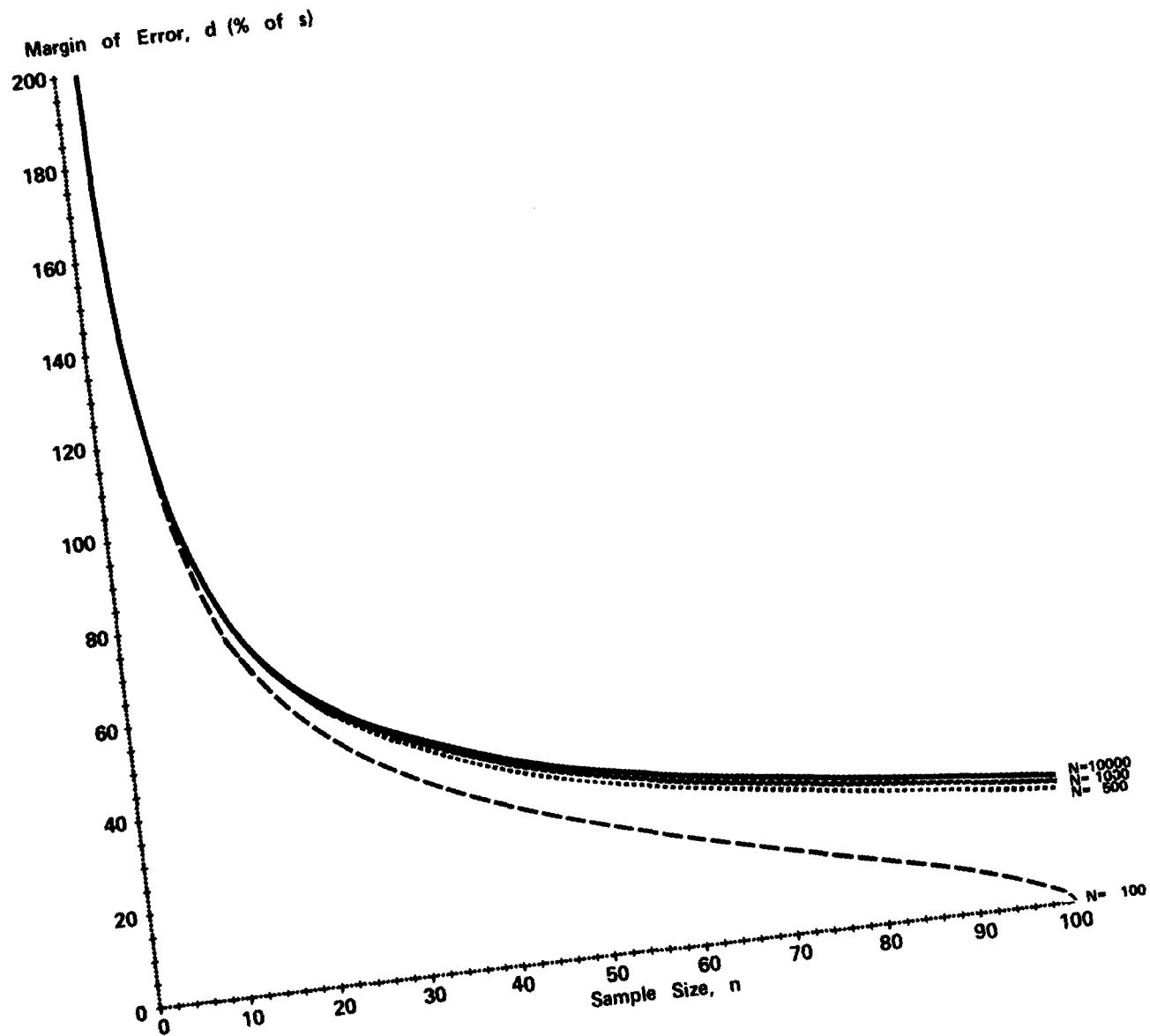


Figure 2. Margin of Error as a Function of Sample Sizes up to 100 in a Finite Population.

The most commonly encountered index of diversity is the Shannon diversity index,¹ d .

Further information about the diversity can be obtained by calculating the relative maximum value of d (d_{max})¹ for that community (which occurs when the number of individuals is equal for all taxa); the relative minimum value of d (d_{min})¹ for that community (which occurs when all taxa except one contain one individual each and the remaining taxon contains the remaining individuals); the efficiency or evenness¹ of the distribution of taxa within that community (the ratio of d_{min} to d_{max}) and the redundancy¹ of that community (1-evenness). These four statistics also apply to the Simpson diversity index,¹ S and the McIntosh diversity index,¹ U .

The above indexes apply to the community as a whole. The following three statistics are applied to each taxon in the community or communities sampled.

The first is relative abundance¹, defined as the number of individuals in a taxon over the total number of individuals in the sample x 100. Those species with a high relative abundance are considered the dominant species in the community.

The second is absolute abundance¹ defined as the number of individuals in a given taxon in a given community over the total number of individuals in all communities sampled x 100. A high absolute abundance identifies that species as rather unique to that community.

The third is the niche breadth,² defined as the measure of how uniformly a given taxon is distributed among the communities sampled.

¹U.S. Army Chemical Systems Laboratory, APG, MD 21010. A Computer Program for the Analysis of Macroinvertebrate Data from Water Quality Surveys, SP4 Cimba, P.A., et al., EO-TR-76102, May 1977.

²U.S. Army Chemical Systems Laboratory, APG, MD 21010. A Revised Computer Program for the Analysis of Macroinvertebrate Data from Water Quality Surveys, Asaki, A.E., et al., In Press.

The basic method used for the analysis of biotic similarity among communities is that of "cluster analysis". To begin with, comparisons are made between all possible community pairs. The Pinkham-Pearson index of biotic similarity, $B^{1,2}$, can be used to compare one pair of communities at a time. Simply stated, B is the sum of the ratios of the minimum value over the maximum value for each species present in one or both of the communities being compared.

The result of application of the index of biotic similarity to all possible community pairs is a matrix of the B 's. The matrix is then used to produce the actual clustering of the resulting similarity values, starting with the highest value of B in the matrix. The results of the cluster analysis can be visually displayed in the form of a dendrogram¹ which resembles a tree on its side. The topmost branches represent those communities with the highest biotic similarity. The final step in the analysis of biotic similarity among communities is the calculation of the cophentic correlation coefficient, r_{sc}^3 , between the original similarity values and those derived from the dendrogram. The cophentic correlation coefficient provides a measure of the distortion introduced by the clustering method.

¹U.S. Army Chemical Systems Laboratory, APG, MD 21010. A Computer Program for the Calculations of Measures of Biotic Similarity Between Samples and the Plotting of the Relationship Between These Measures, Pinkham, C.F.A. et al., AD A007801/4GI, Apr 1975.

²Pinkham, C.F.A. and J.G. Pearson, "Applications of a New Coefficient of Similarity to Pollution Surveys," Journal Water Pollution Control Federation, 48:717-723, Apr 1976.

³U.S. Army Dugway Proving Ground, Dugway, UT 84022. A Computer Program for the Calculations of Cophentic Correlation Coefficients Using the Index of Biotic Similarity, Pinkham, C.F.A. et al, In Press.

APPENDIX A - LIST OF TESTS

<u>Test</u>	<u>Page</u>
Random selection	11
Student's t-test	12
Snedcor's F-test	13
Student-Newman-Keul's test	13
Heterogeneity G-test	13
Simple linear regression	14
Weighted regression	14
Probit analysis	15
Product-moment correlation coefficient	15
Spearman's rank correlation coefficient	16
Multiple regression	16
Multiple correlation coefficient	16
nth-order polynomial	16
Isopleths	17
Stepwise multiple regression	17
Analysis of covariance	18
Hotelling's T ² -statistic	18
Canonical correlation coefficient	18
Multivariate analysis of variance	18
Multivariate regression	18
Multivariate analysis of covariance	18
Principal components analysis	19

LIST OF TESTS (CONTINUED)

<u>Test</u>	<u>Page</u>
Factor analysis	19
Discriminant analysis	19
Shannon diversity index	23
Efficiency	23
Evenness	23
Redundancy	23
Simpson diversity index	23
McIntosh diversity index	23
Relative abundance	23
Absolute abundance	23
Niche breadth	23
Pinkham-Pearson index of biotic similarity	24
Cophenetic correlation coefficient	24

APPENDIX B - DISTRIBUTION LIST

	<u>Number of Copies</u>
Office, Secretary of Defense ATTN: OASD (MRA&L) Rm 3E784, Pentagon Washington, DC 20301	1
Office, Secretary of the Army ATTN: OSA (CW) Rm 2E569, Pentagon Washington, DC 20310	1
Headquarters, Department of the Army Deputy for Science and Technology Office, Assistant Secretary of the Army (R&D) Rm 3E390, Pentagon Washington, DC 20301	1
Headquarters, Department of the Army Office, Chief of Engineers ATTN: DAEN-ZCE Rm IE676, Pentagon Washington, DC 20314	1
Commander US Army Materiel Development & Readiness Command ATTN: DRCIS-A 5001 Eisenhower Avenue Alexandria, VA 22333	2
Commander US Army Test & Evaluation Command ATTN: DRSTE-PP-P Aberdeen Proving Ground, MD 21005	2
Commander US Army Dugway Proving Ground ATTN: STEDP-SD (1) PP (2) MT (1) MT-L (40) MT-DA-L (10) Dugway, UT 84022	54

Number of Copies

Commander US Army Armament Research and Development Command ATTN: DRDAR-PSE Dover, NJ 07801	1
Commander Chemical Systems Laboratory ATTN: DRDAR-CLT-P (1) DRDAR-CLT-E (1) Aberdeen Proving Ground, MD 21010	2
Commander US Army Armament Materiel Readiness Command ATTN: DRSAR-ISE Rock Island, IL 61299	1
Commander US Army Training and Doctrine Command ATTN: ATEN-FE-E Fort Monroe, VA 23651	1
Commander US Army Forces Command ATTN: AFEN-EQ Fort McPherson, GA 30330	1
Commander US Army Health Services Command ATTN: HSL0-F Fort Sam Houston San Antonio, TX 78234	1
Superintendent Academy of Health Sciences ATTN: HSA-IVS Fort Sam Houston San Antonio, TX 78234	1
Commander US Army Engineer Waterways Experiment Station Vicksburg, MS 39180	1

	<u>Number of Copies</u>
US Army Logistics Mgmt. Ctr. ATTN: DRXMC-MR-I Bldg. 12107 Fort Lee, VA 23801	1
Commander US Army Environmental Hygiene Agency Edgewood Area Aberdeen Proving Ground, MD 21010	1
Commander US Army Medical Bioengineering R&D Laboratory Fort Detrick Frederick, MD 21701	1
Commander Construction Engineering Research Laboratory ATTN: CERL-EN P.O. Box 4005 Champaign, IL 61820	1
Commander US Army Toxic and Hazardous Materials Agency ATTN: DRXTH-IR Aberdeen Proving Ground, MD 21010	1
Administrator Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Headquarters, US Air Force ATTN: X00SC (1) LEEV (1) Washington, DC 20330	2
Headquarters AFETO/DEVIN Tyndall AFB, FL 32403	1
Headquarters AFESC/RDVW Tyndall AFB, FL 32403	1

Number of Copies

Headquarters DEOP MacDill AFB, FL 33608	1
Headquarters ADCOM/DEV Peterson AFB Colorado Springs, CO 80914	1
Commander Camp Pendleton Natural Resources Office Camp Pendleton, CA 92055	1
Commander Naval Sea Systems Command ATTN: SEA62R32 Washington, DC 20362	1
Commander Naval Facilities Engineering Command 200 Stovall St Alexandria, VA 22332	1
Atlantic Division Naval Facilities Engineering Command ATTN: Mr. Robert C. Beam Norfolk, VA 23511	1
Northern Division Naval Facilities Engineering Command Bldg. 77, Naval Base Philadelphia, PA 19112	1
Commander Naval Construction Battalion Center ATTN: Navy Environmental Support Office (1) Civil Engineering Laboratory, NCBC (1) Port Hueneme, CA 93043	2

	<u>Number of Copies</u>
Commander Naval Ocean System Center ATTN: Marine Sciences Division San Diego, CA 92152	1
Commander Naval Surface Weapons Center White Oak Silver Spring, MD 20910	1
Commander Naval Air Station ATTN: Public Works Dept. (CODE 6230-2) Port Mugu, CA 93042	1
Executive Office of the President Council on Environmental Quality 722 Jackson Place, NW Washington, DC 20006	1
US Department of Agriculture Plum Island Animal Disease Center Greenport, NY 11944	1
US Department of Agriculture Emergency Programs Staff Hyattsville, MD 20782	1
US Department of Agriculture Veterinary Services Hyattsville, MD 20782	1
US Department of Agriculture Agriculture Research Service Laramie, WY 82071	1
US Department of Agriculture Agriculture Research Service Bee Research Laboratory Tucson, AZ 85719	1
US Department of Agriculture Animal and Plant Health Inspection Service Ames, IA 50010	1

	<u>Number of Copies</u>
US Department of Commerce Environmental Data Service Boulder, CO 80302	1
US Department of Commerce National Marine Fisheries Service Marine Mammal and Endangered Species Division Washington, DC 20235	1
US Environmental Protection Agency Office of Environmental Review 401 M St. SW Washington, DC 20460	1
US Environmental Protection Agency Ecological Research Laboratory Narragansett, RI 02882	1
US Department of the Interior National Fish and Wildlife Laboratory Madison, WI 53706	1
US Department of the Interior Fish and Wildlife Service Patuxent Wildlife Research Center Laurel, MD 20811	1
US Department of the Interior Fish and Wildlife Service Office of Migratory Bird Management Washington, DC 20240	1
US Department of Public Health, Education and Welfare Center for Disease Control Atlanta, GA 30333	1